
MATHEMATICS IN ENGLISH PRINTED BOOKS, 1473–1800:
A BIBLIOMETRIC ANALYSIS

by

BENJAMIN WARDHAUGH*

All Souls College, University of Oxford, High Street, Oxford OX1 4RR, UK

This essay represents a first attempt to make sense of the mass of early modern English publications that deal with or refer to mathematics, using a bibliometric approach made possible by the new electronic databases: *Early English books online* and *Eighteenth-century collections online*. I present statistical information about references to mathematics in this corpus of books, perform some statistical analysis of the trends that the data show, comment on the methodological problems raised, and assess what these results do and do not tell us about early modern English discussion of mathematics.

Keywords: mathematics; books; bibliometrics

INTRODUCTION

A great deal of early modern discussion of mathematics occurred outside works specifically about the subject. Indeed, it seems that something like one-quarter of early English books mentioned mathematics; a total of well over 100 000 works between 1473 and 1800. That proportion changed little, on the whole, between the sixteenth century and the eighteenth, but it fluctuated in the shorter term in ways that hint at a relationship with the political events of the time. The word ‘mathematics’ itself increased in frequency during the period, while ‘astronomy’ decreased (and was briefly eclipsed by ‘astrology’).

This essay is a first attempt to describe, inevitably rather crudely, this corpus of early modern English works that mentioned mathematics; it uses a bibliometric approach made possible by the new digital full-text databases: *Early English books online* and *Eighteenth-century collections online*. I present statistical information about references to mathematics in these databases, perform some analysis of the trends that the data show, and comment on the limitations of this approach and the future lines of enquiry that it suggests. The availability of digital full-text resources of this kind is an innovation of the past five years or so, and it is to be hoped that future work based on them will investigate a wider range of subject areas and employ more sophisticated methods than are used here; the mining of general-purpose resources such as these may—and, I hope, will—prove particularly fruitful for historians of science.

*benjamin.wardhaugh@all-souls.ox.ac.uk

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsnr.2008.0033> or via <http://rsnr.royalsocietypublishing.org>.

METHOD

The catalogues

Two online databases were used in this research: *Early English books online* <eebo.chadwyck.com> (EEBO) and *Eighteenth-century collections online* <galenet.galegroup.com/servlet/ECCO> (ECCO). The work was conducted in April 2008, and what follows is based on the contents of the databases (and the descriptions of them available online) at that time.

The EEBO catalogue, covering the period 1473–1700, is based on two printed catalogues of early English books.¹ An indication of their scope is given by the description of their online successor, *The English short-title catalogue* <estc.bl.uk> (ESTC): it covers ‘letterpress books, pamphlets, newspapers, serials, and a variety of ephemera’, including advertisements; it excludes ‘engraved music, maps and prints, ... printed forms intended to be completed in manuscript, ... trade cards, labels, invitations, bookplates, currency, ... playbills, concert and theatre programmes, ... playing cards, games, puzzles’; it includes material in any language ‘printed in the British Isles, Colonial America, United States of America (1776–1800), Canada, or territories governed by England or Britain before 1801’, or falsely claiming publication in London; finally, it includes material, wherever printed, that is ‘wholly or partly in English or other British vernaculars’.²

EEBO contains nearly 119 000 item records, certainly a high proportion of the extant material within its scope, although that number does not correspond in any easy way to the number of distinct works or editions that the records describe. For about 90% of the items listed it provides digitized images of each page. For a smaller proportion (about 15 000, or about 13%), searchable full text is given, the result of ‘keyboarding’ by human typists.³ The works given as full text are selected on the basis of three criteria: (1) works are included that are present in the *New Cambridge bibliography of English literature*;⁴ (2) second and subsequent editions are usually omitted; (3) extra works are included in response to suggestions and expert advice.⁵

The second database used, ECCO, catalogues ‘138,000 English-language titles and editions published between 1701 and 1800’, a selection from the ESTC—which has nearly 342 000 records for this period—made with the aim of representing specific authors, groups of authors, and subject areas. No detailed information is given, but the selection has apparently been guided by a categorization of items into seven subject areas: history and geography; social science and fine arts; medicine, science and technology; literature and language; philosophy and religion; law; and general reference. The sheer size of the selection suggests that in many cases it is editions rather than works that are omitted. Here every item is supplied with both digitized page images and full text, the latter originating from an optical character recognition process that provides rather poorer accuracy than the EEBO full texts.⁶

As far as available full texts are concerned, EEBO provides a selection of about 13% of the items within its scope, and ECCO about 40%; this, together with the differences in their selection policies and in the accuracy of the full texts they provide, is an important difference between the two databases for this research.

These resources are widely used by historians, but they are by no means immune to the problems common to all catalogues—error, omission, duplication, mistaken inclusion⁷—to which I will return in the final section of this paper. This paper attempts to find out what can be done with these resources despite these problems, and sketch a methodology that, I hope, can in the future be made more sophisticated and robust by others.

The two databases exclude journals and other periodicals. Because journal publication raises such distinctive issues for the use of mathematical vocabulary, this limitation need not be considered a problem; however, it should be borne in mind. They also, obviously, exclude material that circulated in manuscript only. Correspondence—sometimes circulating very widely—was the vehicle for a great deal of early modern technical writing, including mathematics, and scribal ‘publication’ continued, even thrived, well into the eighteenth century, so that certain literary forms are substantially under-represented here: a great deal of early modern poetry, for example, circulated only in manuscript,⁸ and the metaphysical tendency of the late sixteenth and early seventeenth centuries resulted in many poetic references to mathematics that would be relevant here (consider the famous ‘compass’ figure in John Donne’s ‘A Valediction: forbidding mourning’, for example).

The searches

For each decade from 1481–90 to 1691–1700, and for the part-decade 1473–80, EEBO was searched seven times, as follows:

1. with a blank search string and a restriction to items for which full text was available;
2. with full-text search strings of ‘mathemati*’, ‘geomet*’, ‘arithmeti*’, ‘astro* OR afro*’, and ‘algebr*’;
3. with a full-text search string of ‘mathemati* OR geomet* OR arithmeti* OR astro* OR afro* OR algebr*’.

For each decade from 1701–10 to 1791–1800 ECCO was searched with a blank search string, to produce a count of the items in the catalogue, and with the six full-text search strings listed above.

The choice of search terms was of course of the first importance, and the terms were chosen, as far as possible, to match cognate words and variant spellings: ‘mathematician’, ‘mathematicall’, ‘mathematicks’ or ‘mathematique’, for example. The search term ‘astro*’ matches both ‘astronomy’ and ‘astrology’ and their cognates (a brief attempt to disentangle the two is given below). The term ‘afro*’ was included to catch the (frequent) case of the long s being misread by ECCO’s character-recognition software.

The works searched were in every case restricted to those in English, both because EEBO provides very few full texts in other languages and to avoid the danger of the search terms’ producing unwanted results. A study of the Latin publications which are enormously important for early modern mathematics would require a rather different approach from this one, because of their international circulation.

RESULTS

Numbers of books

The mean number of works that mention any of these items of mathematical vocabulary is 28.7% of the total; 17.5% mention mathematics itself, 14.7% astronomy/astrology, 10.7% and 10.4% respectively geometry and arithmetic, and 3.7% algebra.⁹ These numbers are quite high, and the fact that more than one-quarter of early English books apparently mentioned mathematics in one way or another is a striking one. Equally striking is the fact that the word ‘mathematics’ was mentioned by more works than astronomy, geometry or arithmetic. These figures alone point to a whole world of discussion of mathematics—even if it consisted

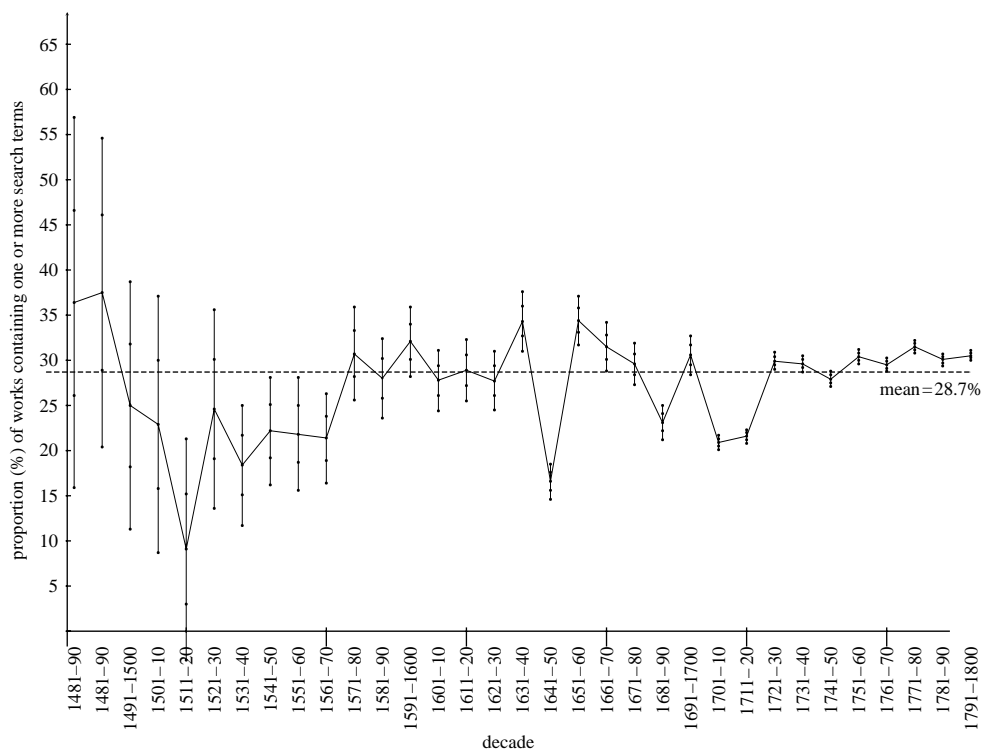


Figure 1. Proportions of works that mention one or more of the search terms.

only of passing references—which is invisible in our histories of mathematics: 28.7% of the English books printed by 1800 would be rather more than 100 000.

In fact these numbers are somewhat misleading, because the proportions vary quite substantially from one decade to another. Some of the terms show clear long-term trends in their frequency, but others do not. Figure 1 shows this variation, decade by decade, in the proportion of works that mention any one (or more) of my search terms. To it are added error bars set at two standard errors,¹⁰ representing the range within which, with 95% confidence, we can say that the ‘true’ proportion—the probability of an author’s¹¹ mentioning mathematics in his or her book—lay for each decade. Because the sample size—the number of books available for searching—increases for later decades (from just 22 to more than 28 000, in fact), the error bars become much smaller towards the right of the graph.

This image illustrates that the proportion of books mentioning mathematics fluctuated from decade to decade. More importantly, it shows that some of those fluctuations take the proportion more than two standard errors from its mean of 28.7%: the dips in 1511–20, 1531–70, 1641–50, 1681–90 and 1701–20, for example. They are therefore ‘real’, in the sense that they cannot be explained by supposing that an author’s probability of mentioning mathematics in his or her book was constant between 1473 and 1800, and that the variations we observe are the results of chance.

More precisely, a χ^2 test allows us to test the hypothesis of an underlying constant mean for these data, and it shows that that hypothesis must be rejected on the basis of these data as they stand.¹² By eliminating from consideration, one by one, those decades whose

proportions (weighted for their sample size) deviate most from the mean, we can, eventually, arrive at a reduced data set for which we can accept the hypothesis of a constant underlying mean. The decades that must be eliminated are, in order, the 1710s, 1700s, 1640s, 1680s, 1740s, 1770s, 1650s, 1560s, 1530s, 1630s, 1540s, 1550s and 1510s.¹³

Are there plausible historical reasons why these decades may have been exceptional in this particular respect? Here we are on dubious ground. It is striking that the first five decades on my list of exceptions correspond to periods of war or political upheaval for Britain: the War of the Spanish Succession (1701–14), the Civil War(s) (1638–51), the ‘Glorious’ Revolution (1688–89) and the War of the Austrian Succession (1740–48). It is equally obvious that such post hoc correspondences could be found for any set of supposedly exceptional decades. For the decades further down the list we might have to work a little harder: whereas the sixteenth century provides any number of disruptive events that might be linked with the troughs of 1511–20 and 1531–70, for the elevated proportions for the 1630s, 1650s and 1770s no obvious correlations spring to mind. Conversely, one could identify any number of clearly important events from this period for which the statistics show no corresponding peak or trough: the Anglo-Dutch wars of 1652–54, 1665–67, 1672–74 and 1780–84; the plagues of 1603, 1625, 1636 and 1665; the fire of 1666, for example. In any event, correlation is not causation, and what appear in these statistics as dips and troughs may equally well be the result of a fairly flat rate of ‘mathematical’ publication being, in some decades, overshadowed by an inflated rate of publication (or survival, or cataloguing) of, say, political pamphlets.

To interpret these variations, it would be helpful to have similar statistics for a different set of search terms—a group of medical terms, or political ones, for example—for comparison. This would at least help us to know whether a historical explanation specific to mathematics is needed.

That said, I offer [figure 2](#) without further comment. Based on statistics for each year of the period, it displays for each year the average proportion found over the 11 surrounding years: at the expense of some artificiality, this shows a reasonably smooth curve to which certain political events—wars and changes of ruler—can be correlated although others cannot. What this means, or indeed whether it means anything, I am unable to say.

Rather than identify half of the sixteenth century as being in some unspecifiable sense exceptional, it may be more helpful to observe that the mean for the sixteenth century (26.7%) is somewhat lower than that for the eighteenth (28.8%), and that in the seventeenth century it lies between the two (27.9%). It is tempting, indeed, to attempt to fit a slope or a curve to the data rather than a flat mean; however, [figure 1](#) does not really fit any smoothly rising curve unless a large number of decades are ignored, and—after experimentation and with some regret—I have concluded that the use of regression techniques, at least at my own limited level of statistical expertise, does not tell us anything of substance about these data that we cannot discern by inspecting a graph. Apart from any other consideration, it is not at all clear that linear or logistic relationships are ones we would plausibly expect to find in the historical development of a quantity like this one. (It may also be mentioned here that a scatter plot of the proportion for each decade against the proportion for the previous decade—seeking evidence of ‘autocorrelation’, the effect that one decade’s proportion is a function of the previous decade’s—shows no discernible pattern of any kind.)

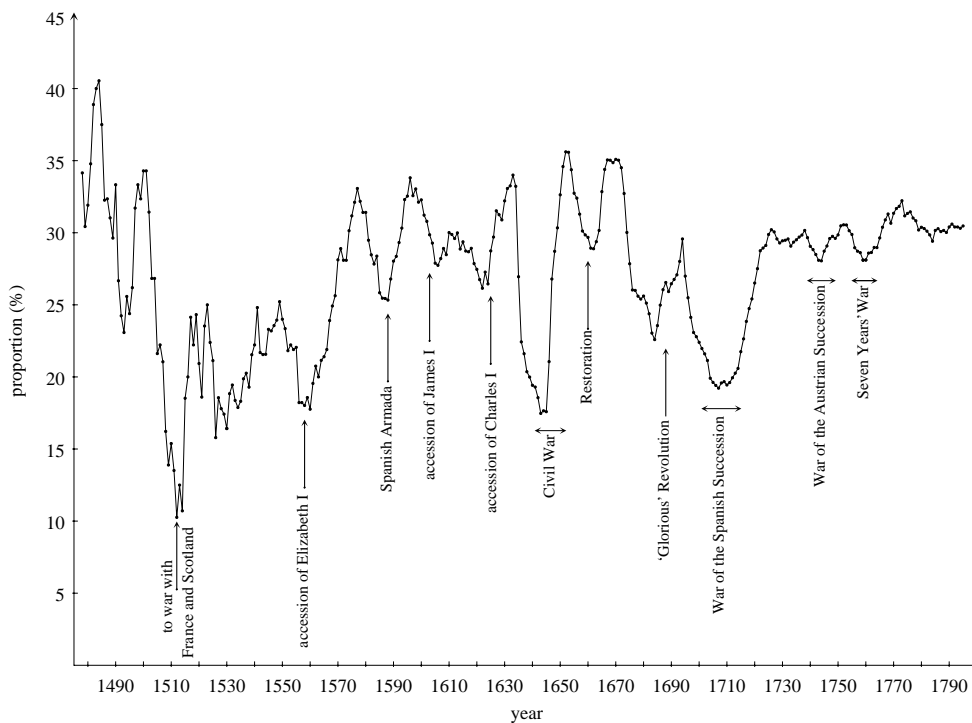


Figure 2. Proportion of works that mention one or more search terms: 11-year averages.

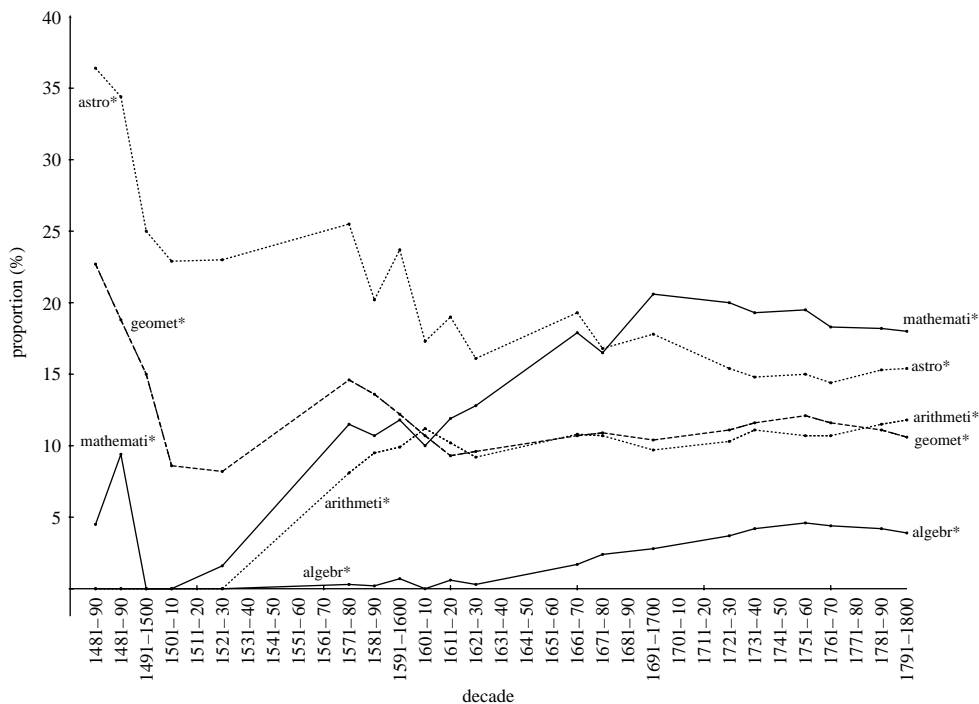


Figure 3. Proportion of works that mention individual search terms.

Table 1. Mean proportion of works mentioning each search term, by century.

century	percentage of works mentioning					
	mathemati*	geomet*	arithmeti*	astro*	algebr*	any
sixteenth	9.5	11.0	6.9	20.4	0.5	26.7
seventeenth	14.9	9.7	9.7	17.1	1.5	27.9
eighteenth	17.9	10.8	10.5	14.4	3.9	28.8

Individual search terms

Similar remarks can be made for the individual search terms. Each settles at a stable proportion towards the end of the period, and ‘geometry’ remains very roughly constant around its mean of 10.7% throughout the entire period. ‘Arithmetic’, not found at all before 1520, rises from nothing to reach around 10% by 1580, where, roughly, it remains. ‘Algebra’, also not found before 1520, rises to a stable value around 4% by about 1730. ‘Mathematics’, found at very low values before 1510, rises to around 18% by 1690. The single falling trend is shown by ‘astronomy/astrology’, which falls to around 15% by 1720, having shown values of 25% or higher in the sixteenth century.

All of these trends are shown in [figure 3](#), from which the decades identified above as exceptional are excluded to make the trends clearer. It is not difficult to fit, by eye, reasonably plausible best-fit curves to these data, but any attempt at a more sophisticated statistical analysis founders on the quite substantial variability of all these proportions from decade to decade, with or without the ‘exceptional’ decades, and the fact that all seem to have periods of constancy as well as periods of change. What we can say, however, is that the differences that the proportions show from one century to another are certainly large enough to be statistically significant.¹⁴ These century-wide proportions are shown in [table 1](#).

It would be difficult to exhaust the possibilities of these data in a short paper, and readers who are interested are welcome to perform analyses of their own on the raw data which are given in appendix 1 in the electronic supplementary material. I will look briefly at two questions and one possible objection.

Bias in ECCO

First, the objection. Does ECCO artificially reduce the numbers for the eighteenth century by using a relatively inaccurate optical character recognition process? A comparison of the results for the 1690s (EEBO) and the 1700s (ECCO), in isolation, might suggest so: most of the proportions jump downwards, some of them quite sharply, between the two. But by the 1720s they have recovered to the levels of the 1690s, and the 1700s were identified above as an exceptional decade in relation to the whole of the data—which suggests that the jump is real rather than an artefact.

Alternatively, does ECCO bias the statistics by recognizing some words with significantly greater accuracy than others? One way to test this is to check whether the *relative* frequencies of the different search terms seem to change sharply at the transition between the two databases. In brief, they do not (see [table 2](#), which gives the number of hits for, say, ‘mathematics’ as a proportion of the total number of hits for all the search terms): although there are changes, they are within the range that we see elsewhere in the data, and a graph of these ‘relative’ proportions does not show unusual changes at the transition between the databases, which would suggest a problem.

Table 2. Numbers of hits for each search term as a proportion of the total number of hits, 1681–1720.

decade	number and percentage of hits					total
	mathemati*	geomet*	arithmeti*	astro*	algebr*	
1681–1690	251 (55%)	143 (31%)	150 (33%)	276 (60%)	37 (8.1%)	457
1691–1700	379 (67%)	192 (34%)	179 (32%)	329 (58%)	52 (9.2%)	564
1701–1710	1253 (60%)	727 (35%)	723 (35%)	1078 (52%)	187 (9.0%)	2077
1711–1720	1461 (60%)	809 (33%)	821 (34%)	1141 (47%)	214 (8.8%)	2434

This is, unfortunately, probably as much as we can say about ECCO's accuracy. Any attempt to test it more directly, using manual checking, would need to take into account that it must depend on factors that vary from work to work (the quality of the type, the condition of the copy used, the frequency of italics, for instance): I performed a very brief spot check and found success rates as low as 33% and as high as 93% for the recognition of individual occurrences of 'arith*' in different works. The construction of a completely representative sample of works to check by hand, even for a single decade, seems scarcely feasible. Of course, ECCO's success at identifying *works* that contain a particular term also depends on the frequency of occurrences of the term per work (spot checks suggest that the modal frequency is typically 1, and the mean often around 2).

Thus, although it does not, informally, really look as though ECCO biases the numbers in either of these ways, we cannot be certain of that. We consequently cannot be certain that perceived similarities or differences between the populations of printed books in the two periods covered by the two databases are not artefacts of the differences between the databases' search functions.

Astronomy and astrology

Next, the question of 'astronomy' versus 'astrology'. Perusal of early modern works that use them, and particularly the many almanacs of the period, indicates that the relationship between the two terms changes over time: if they are seldom synonyms they are certainly not always distinguished with much clarity, and neither has a meaning that remains wholly constant. But it is difficult to be specific about the nature or even the direction of any shift between the two. This statistical approach therefore provides an opportunity to specify at least the quantitative element of the relationship between the two words: which was used more frequently, and how does the answer to that question vary over time?

This can be answered by using searches for the terms 'astronom*' OR 'aftronom*', and 'astrolog*' OR 'aftrolog*'. (These longer search terms, as well as distinguishing between 'astronomy' and 'astrology', attempted to ensure that peripheral terms such as 'astrolabe' or 'astroid' (i.e. asteroid) were omitted.) Overall, 'astronomy' was the more frequent term, mentioned in 10% of the works searched, in comparison with 6% for 'astrology'. Quite a number of works—3% of the total—mentioned both.

When the data are viewed decade by decade, the remarkable feature emerges that 'astronomy' both starts and ends the period as the more frequent term, but—relative to 'astrology'—it shows quite a marked fall from the early sixteenth century until the 1660s, followed by an equally marked recovery. 'Astrology' shows the opposite pattern, steadily increasing in frequency to the middle of the seventeenth century and declining thereafter. The result is that between about 1650 and 1720 'astrology' is the more frequent term. All of this

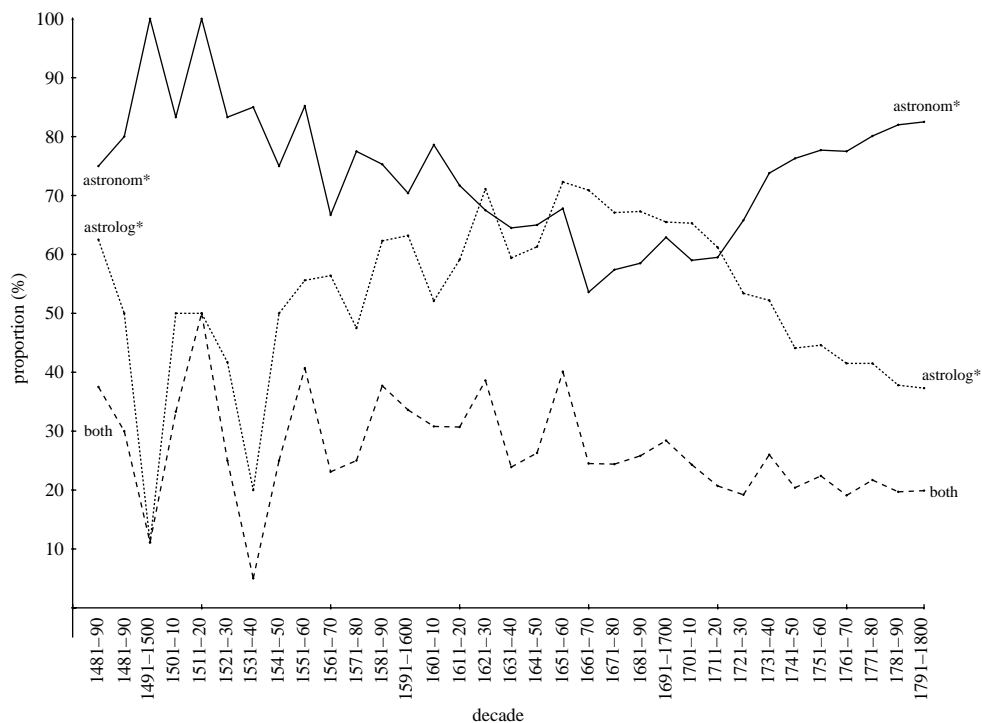


Figure 4. 'Astronomy' and 'astrology': relative proportions.

takes place within the general pattern noted above, that both terms become less frequent across the whole period. This is shown in figure 4, in which the number of works mentioning 'astronomy' or 'astrology' is plotted as a proportion of the number of works that mention either term. (Raw data are given in appendix 1 in the electronic supplementary material.) Also shown in the graph is the proportion that mention *both* terms.

These results are about the frequency of the terms in the sense of the number of works that use them, even in passing. A work that contains a single disparaging reference to astrology looks exactly the same in these statistics as one that praises the subject on every page, so that we cannot reason from frequency to popularity or prestige.

Overlapping terms

Finally, the technique of considering the number of hits for one search term as a proportion of the number of hits for any search term lends itself to a brief extension. Nearly every work that mentioned any of the terms mentioned 'astronomy/astrology' in the early sixteenth century, for example, whereas for the whole of the eighteenth century between 60% and 70% mentioned 'mathematics', around 50% 'astronomy', and 35–40% 'arithmetic' or 'geometry'. There must therefore have been substantial overlap between the works that used the different terms. In fact, a work that contained any one of the five search terms contained, on average, about two of them. But which two? Which combinations occurred most frequently?

This is answered by searches for terms such as 'mathemati* OR geomet*', for each of the three centuries from 1501. In the sixteenth century a work that mentioned 'mathematics', 'geometry' or 'arithmetic' usually (around 70% of the time) also mentioned 'astronomy', and

Table 3. The overlap of search terms.

	percentage overlap, given...				
	mathematics	geometry	arithmetic	astronomy	algebra
1501–1600					
mathematics	–	45	47	33	80
geometry	52	–	70	38	100
arithmetic	34	44	–	23	80
astronomy	70	71	68	–	90
algebra	4	4	5	2	–
1601–1700					
mathematics	–	66	55	48	96
geometry	43	–	52	37	87
arithmetic	36	52	–	32	73
astronomy	55	64	57	–	81
algebra	10	14	12	7	–
1701–1800					
mathematics	–	72	61	52	84
geometry	43	–	50	38	74
arithmetic	36	49	–	31	70
astronomy	42	51	42	–	64
algebra	18	27	26	17	–

a work that mentioned ‘arithmetic’ usually (again, around 70% of the time) also mentioned ‘geometry’. No other association was as strong as these. In the seventeenth century the strongest associations between search terms were that works mentioning ‘geometry’ usually also mentioned ‘mathematics’ or ‘astronomy’ (each around 65% of the time, so plenty of works must have mentioned all three). In the eighteenth century the pattern changed again: 72% of works mentioning ‘geometry’ also mentioned ‘mathematics’, as did 61% of works mentioning ‘arithmetic’. Throughout the three centuries, ‘algebra’ rarely appeared alone, and its most frequent companion in the seventeenth and eighteenth centuries was ‘mathematics’.

This hints that in the sixteenth century authors mentioning mathematical topics tended to link them to astronomy; in the eighteenth century they tended to link them to ‘mathematics’ itself, and in the seventeenth century there was rather more general overlap between the use of the four terms, with ‘geometry’ possibly the term most often linked with another. Table 3 displays all of the proportions for these overlaps.

DISCUSSION

Some of what has been said accords with what is well known to historians of the mathematics of this period. Algebra, for instance, was becoming more widely used, and the term ‘algebra’ was (slowly) gaining ground over alternatives such as ‘the analytic art’ or ‘the cossic(k) art’; and thus we see the frequency of the term increasing in print throughout much of this period. However, algebra remained a minority practice, whereas the period saw an increasing tide of publications aimed at readers from the lower and middle classes who wished to learn or improve numerical skills for use in work or trade: thus, the terms ‘arithmetic’, ‘geometry’ and ‘mathematics’ show increases, to much higher frequencies than ‘algebra’, during the period. If my results do no more than to add a quantitative dimension to our knowledge of such trends they have, I think, served a worthwhile purpose.

They do, however, do more, and I would like to highlight three points before moving on to discuss the (many) limitations of this approach. The first is the sheer number of works that are identified as mentioning one or more of my mathematical search terms, a number that cannot be accounted for by works that are in any sense ‘about’ the terms in question: more than 3000 works mentioned arithmetic in the 1790s, for example, and more than 1000 mentioned algebra, corresponding to possibly twice that many in the total corpus of works from which ECCO provides a selection. These numbers indicate, rather, that mathematical terms were very often mentioned in works in which they were not central concerns. Mathematical terms could serve as metaphors; they could feature in quasi-mathematical methods of proof; they could be used casually or negatively. (This last point is important: to use a term is not to be enthusiastic about either it or what it stands for.) To investigate these uses, and their relationship to writings actually *about* mathematics—concerning which these statistics tell us nothing—is a task for which entirely different methods are needed.

Second, as we have seen, the terms ‘astronomy/astrology’ did not just decline in frequency over the period, as we might deduce from a consideration of the declining currency of the astrological prognostications found in cheap almanacs at least from the mid seventeenth century onwards.¹⁵ There was a striking evolution in which the terms exchanged their relative frequencies for part of the seventeenth century. This, too, suggests a line of enquiry into just what those terms meant during that unusual period, and how that differed from their meanings at other times.

Finally, we have seen that certain decades are to be identified as exceptional in the frequency with which mathematical vocabulary appeared in printed books: in most of them that vocabulary was unusually rare, beyond what can be accounted for by random variation in the samples we have available for searching. Thus, these exceptional decades were apparently ones of real variation in the currency or relevance of these terms, and I have very tentatively suggested that a correlation with periods of political upheaval—rather than with disruptions of other kinds—might be discerned. Of course, the point of any further work on such correlations would be precisely in the details that a statistical approach ignores: individual cases of delayed or abandoned publication and the circumstances that led to them, for example, or, conceivably, documentable enthusiasm for particular subjects in the wake of prominent printed expositions such as Billingsley’s 1570 *Euclid* or Newton’s 1687 *Principia mathematica*; but also the cases in which unfavourable circumstances did *not* result in a publication’s being abandoned or in which a high-profile exposition did *not* result in a rush of imitators.

LIMITATIONS

These statistics therefore suggest several specific worthwhile questions about early modern discussions of mathematics in print; I hope to pursue some of them in the future. However, before concluding, it is also very important to understand the serious limitations inherent in these data and this method.

These data and statistics describe the full-text corpora of EEBO and ECCO; how far we can reason from them to (surviving) early English books as a whole it is difficult even to guess. The selection of works for full-text presentation in these databases is unlikely to be entirely independent of the works’ genres and subjects, and therefore of whether they are likely to use any mathematical vocabulary. Perhaps for periods of political uncertainty the selectors’ attention turns towards works that focus exclusively on political events; perhaps

EEBO's policy of preferring works whose authors feature in a standard bibliography causes it to over-represent mathematical works, which may have been printed anonymously more rarely than others; perhaps both catalogues' preference for displaying mainly first editions as full text may have caused often-reprinted mathematics primers to receive less emphasis here than they should. One certain example of such a problem is the disfavour in which EEBO apparently holds almanacs: for, say, the 1670s more than 200 are catalogued, but only 3—rather than the 30 or so that one would expect—are given as full text. Problems of this kind are simply unavoidable as long as EEBO and ECCO provide full text for only a selection of the extant works from their respective periods (ECCO's much larger selection of books than EEBO's for full-text presentation may make the statistics for the eighteenth century somewhat more robust in the face of these problems than those for the sixteenth and seventeenth centuries)—unless one were to construct independently a large and genuinely random sample of early English printed works and test it for the characteristics sought here, a task well beyond the bounds of feasibility. Thus, both the long-run picture and the short-term changes I have identified are capable, in principle, of being artefacts of the selection of works that are available as full text.

EEBO's and ECCO's representation of early English books is also affected by the common problems of large bibliographic catalogues: the accidental inclusion of non-existent or irrelevant items, the omission of others (neither database makes it clear what mechanisms exist for correcting such errors if they are detected), and the question of when variant copies should be catalogued as separate impressions or editions.¹⁶ Although these issues are worrying, it is, perhaps, relatively difficult to envisage plausible ways in which they would significantly affect either the proportion of works that mention a particular item of vocabulary, or the variation of such a proportion over time.

If these statistics do not describe 'surviving early English books' but the EEBO and ECCO full-text corpora, still less do they describe the total population of the works that were actually printed in English, many of which have *not* survived. Mathematics primers and practical manuals, often flimsy and intended for practical use, have probably fared badly in this respect. (The effects of non-random survival are probably smaller, though, in databases as comprehensive as these than in a more traditional study based on the perusal of the physical holdings of an individual library.) Nor do the statistics tell us about the population of printed works actually in circulation in any early modern year or decade: the unit for this research has been the work, not the copy, and the variable—and very often unknown—sizes of early modern print runs make it impossible to reason from one to the other. Nor, of course, are these statistics a description of the attitudes of the English-literate population, or of any individual. If there were many locations for the discussion of mathematics in printed books and pamphlets, there were many more elsewhere, not least in the irrecoverable spoken word.

Finally, the fact that my search terms (deliberately) conflate pairs such as 'mathematics' and 'mathematician' makes this method a decidedly blunt instrument; one that, although I believe it can tell us something about the prominence of mathematics, geometry, and the rest, in early modern English print culture, cannot be used to reason about their prestige or their status. Limited to a small set of search terms, these searches do not pretend to capture the whole of the range of mathematical discussion, in which mathematics might be used without being named at all (the implicit presence of mathematical calculation in certain types of political or economic writing, for example, would be invisible to this or any similar technique) or, conversely, may be absent from a work that happens to use such a rare term as 'astroid' (asteroid: and what similar oddities may lurk unnoticed in the results of my

searches?). And, as I mentioned above, these searches take no account of the number of references to mathematics found in each work, and thus—perhaps their single most important limitation—do not distinguish between works that are about mathematics, those that refer to it in passing (whether positively or negatively), or works such as encyclopaedias, which have a section about mathematics. (A study of mathematical vocabulary on title pages would provide a valuable comparison in this respect.)

But these statistics do tell us several striking and novel things about mathematics in early modern English books. Perhaps their main function is to raise questions: What kinds of works mentioned these words? How frequently did they mention them, and what uses did they make of them? What kinds of authors or audiences were involved? And, of course, what can all this tell us about the status and the meaning of mathematics across this period? I hope that my results will provide a stimulus to work on these questions, as well as to statistical studies of other vocabularies.

ACKNOWLEDGEMENTS

I am grateful to Jacqueline Stedall, Will Poole and two anonymous referees for their insightful comments on earlier versions of this paper, and to the members of the History of Mathematics seminar at the Queen's College, Oxford, for their encouragement of the earlier stages of this research.

NOTES

- 1 A. W. Pollard and G. R. Redgrave, *A short-title catalogue of books printed in England, Scotland and Ireland and English books printed abroad, 1475–1640* (Bibliographical Society, London, 1926; revised second edition 1976–91); Donald Wing, *Short-title catalogue of books printed in England, Scotland and Ireland, Wales and British America and of English books printed in other countries, 1641–1700* (printed for the Index Society by Columbia University Press, New York, 1946–51; revised second edition published by the Modern Language Association of America, 1982–98). EEBO records refer to the revised editions of these works, but its catalogue seems to derive ultimately from the catalogue of the Early English Books microfilms project, which began in 1937 (*Early English books, 1475–1640* (UMI, Ann Arbor, 1937–) and *Early English books, 1641–1700* (UMI, Ann Arbor, 1975–)).
- 2 <<http://www.bl.uk/reshelp/findhelprestype/catblhold/estccontent/estccontent.html>>. EEBO also includes the works listed in two other catalogues, the *Thomason tracts (1640–1661)* collection and the *Early English books tract supplement*: these supplementary collections were excluded from this research. <<http://eebo.chadwyck.com/about/about.htm>>.
- 3 <<http://www.lib.umich.edu/tcp/>>. Strictly, the Text Creation Project (TCP), run from the University of Michigan, is a separate project from EEBO and functions autonomously; in fact, because the EEBO web interface provides ready access to the TCP texts, updated daily, it is convenient simply to refer to these as part of the EEBO data set.
- 4 George Watson, I. R. Willison and J. D. Pickles, *The new Cambridge bibliography of English literature* (Cambridge University Press, 1969–77).
- 5 <http://www.lib.umich.edu/tcp/eebo/proj_stat/ps_text.html>.
- 6 <<http://galenet.galegroup.com/servlet/ECCO>>, following links to 'help' and 'about'. (ECCO's page URLs are valid only within the institution from which they are accessed, and are not given here.)

- 7 See David McKitterick, “‘Not in STC’: opportunities and challenges in the ESTC”, *Library* **6**, 178–194 (2005).
- 8 Consider the Perdita project, for example, which provides a catalogue of early modern women’s manuscripts, many of which include poetry otherwise unknown: <<http://human.ntu.ac.uk/research/perdita>>.
- 9 To avoid clogging this paper, I will take ‘and its cognates’ to be understood from here onwards.
- 10 One standard error is $s\sqrt{p/(1-p)}$, where s and p are the sample size and proportion for the decade, respectively.
- 11 Strictly, an author–publisher–printer team, whose interrelations could be complex: ‘author’ is a shorthand.
- 12 $\chi^2=485.6$ with $N=33$; $p<0.01$.
- 13 This allows us to accept the hypothesis of a constant mean of 30.1% at the 5% confidence level: $\chi^2=8.87$ with $N=20$; $p>0.95$. If the 1790s and 1620s are also eliminated we can accept a constant mean of 29.9% at the 1% level: $\chi^2=6.22$ with $N=18$; $p>0.99$.
- 14 Specifically, using a z -test for equality of proportions, all the changes between the sixteenth and seventeenth centuries—including that in the overall proportion mentioned above—are significant at the 1% level except that for ‘geometry’. Between the seventeenth and eighteenth centuries all but ‘arithmetic’ are significant at the 1% level: and the latter is significant at the 5% level.
- 15 See Bernard Capp, *English almanacs, 1500–1800: astrology and the popular press* (Cornell University Press, Ithaca, NY, 1979), *passim*; Cyprian Blagden, ‘The distribution of almanacks in the second half of the seventeenth century’, *Stud. Bibliogr.* **11**, 108–117 (1958).
- 16 See Stephen Tabor, ‘ESTC and the bibliographical community’, *Library* **8**, 367–386 (2007); the correspondence of Peter W. M. Blayney, Henry L. Snyder and M. J. Crump in *Library* **1**, 72–78 (2000); Peter W. M. Blayney, ‘The numbers game: appraising the revised *Short-title catalogue*’, *Pap. Bibliogr. Soc. Am.* **88**, 353–407 (1994); Peter W. M. Blayney, ‘STC publication statistics: some caveats’, *Library* **8**, 387–397 (2007). The last of these provides an extremely salutary warning for all who attempt statistical work of this kind.